

## HOW TO LINK CPS PUBLIC USE DATA FILES

This documents the mechanics involved in linking CPS data files. There are several scenarios, but in each case there are three basic steps:

1. Create data extracts from each of the files to be linked;
2. Sort the files in common order by using the variables identified as match keys; and
3. Merge the files.

The March Supplement, also known as the ASEC, file is structured differently than non-March Supplement data files, so the identifiers are not the same.

### **I. March Supplement/ASEC Files**

There are two basic limitations in linking the March CPS files across years. First, only fifty percent of the sample is included in two consecutive years. Second, the residents within the eligible housing units may have changed or appeared as noninterview records in one or both years. The result is a matched sample of considerably less than the upper limit of fifty percent. The basic procedures and variables used to link two or more March CPS files are given in the table below.

**Sample Selection:** Prior to 2002, take these steps to begin the match. The first step in matching year  $t$  with year  $t+1$  is to select from year  $t$  those housing units with a "month in sample" value of 1 through 4, and from year  $t+1$  those units with a "month in sample" value of 5 through 8. This will identify the sample subset eligible for matching. Within this subset, housing units in year  $t$ , month 1 will match only with units in year  $t+1$ , month 5, month 2 with month 6, and so on.

For 2002 forward, H-MIS cannot be used to match households.

**Matching Housing Units:** Using one or more variables, it is possible to uniquely identify each housing unit in each sample rotation. However, because of changes in CPS procedures, the identifiers are not always the same. Below are the variables available for matching March/ASEC CPS files.

Year	Identifiers		
1994 – 2001	H-MIS	H-IDNUM	
2002 – 2004	H-IDNUM	H-HHNUM	
2005 - present	H-IDNUM1	H-IDNUM2	

### **II. Non-March Supplement Files**

The first limitation mentioned above for the March Supplement Files does not exist for linking non-March Supplement files, because a household is normally in sample for 4 consecutive months in year  $t$ , and then in those same 4 months in year  $t+1$ . Use the "month in sample" variable only when linking housing units across years. Use the identifiers below, associated with the month and year the survey was conducted.

Month & Year	Identifiers			
Jan, 1994 – April 2004 <sup>i</sup>	HRHHID	HRSAMPLE	HRSESRUF	HRHHNUM
May 2004 - present	HRHHID1	HRHHID2		

### Linking persons across data files

If you desire to link not only the household information, but the person data as well, follow the procedure above, but add one or more variables to uniquely identify a person. For non-March files, add PULINENO, PUSEX, AND PRTAGE. Remember that the age will increase by 1 in the year  $t + 1$  file.

For March/ASEC files from 1994 to 2004, add A\_LINENO, A\_SEX, AND A\_AGE. For 2005 forward, one variable may be used by itself instead of adding it to the household identifiers. PERIDNUM is the only identifier needed for linking persons in files from 2005 forward.

### **III. Linking ASEC files to Non-ASEC files**

Sometimes, there's a need to link an ASEC (or March supplement) file to a non-ASEC file. Follow the matchkeys below to match households, pertaining to what year the survey was conducted.

#### FOR THE ASEC FILE

Year	Identifiers		
1994 – 2004	H-IDNUM	H-HHNUM	
2005 - present	H-IDNUM1	H-IDNUM2	

#### FOR THE NON-ASEC FILE

Month & Year			
Jan, 1994 – April 2004 <sup>ii</sup>	HRHHID	HRHHNUM	
May 2004 - present	HRHHID1	HRHHID2	

If you desire to link not only the household information, but the person data as well, follow the procedure above, but add one or more variables to uniquely identify a person. For non-March files, add PULINENO, PUSEX, AND PRTAGE. Remember, however, that the age may differ slightly from one file to the other.

For March/ASEC files from 1994 to 2004, add A\_LINENO, A\_SEX, AND A\_AGE. For 2005 forward, one variable may be used by itself instead of adding it to the household identifiers. PERIDNUM is the only identifier needed for linking persons in files from 2005 forward.

---

<sup>i</sup> For files April 2004 to June 1995, you must add the state code, gestcen, to the list of identifiers to uniquely identify households. Due to the phase-in of the 1990 sample, a small number of households will share identity unless adding this code.

<sup>ii</sup> For files April 2004 to June 1995, you must add the state code, gestcen, to the list of identifiers to uniquely identify households. Due to the phase-in of the 1990 sample, a small number of households will share identity unless adding this code.